dergipark.org.tr/tr/pub/19maysbd

19 MAYIS JOURNAL OF SOCIAL SCIENCES

19 MAYIS SOSYAL BİLİMLER DERGİSİ

ISSN: 2717-736X

Cilt 5 Sayı 3

DOI: 10.52835/19maysbd.1582285

Gönderim Tarihi	**09.11.2024**
Kabul Tarihi	**27.11.2024**

Joint Correspondence Analysis: A Case Study on Life Satisfaction in Türkiye

Bileşik Uygunluk Analizi: Türkiye'deki Yaşam Memnuniyeti Üzerine Bir Uygulama

Elif ÇOKER*

* Dr. Öğr. Üyesi, Mimar Sinan Güzel Sanatlar Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, elif.coker@msgsu.edu.tr, ORCID:0000-0003-2572-3654

Abstract

Correspondence analysis is used both as a numerical and mostly as a graphical analysis when the data set is expressed as tables that contain frequencies or counts. It enables researchers to comprehend and interpret the relationship between the variables included in a contingency table. Simple correspondence analysis explores the relationship between two variables in a contingency table. However, most practical situations involve more than two variables, for which multiple correspondence analysis is employed to investigate the relationships among them. Joint correspondence analysis is proposed as an improvement of the multiple correspondence analysis. The aim of this study is to analyze and interpret the data from the survey of life satisfaction in Türkiye through joint correspondence analysis and alternative multiple correspondence analysis.

Keywords: Joint correspondence analysis, Multiple correspondence analysis, Life satisfaction survey

Öz

Uygunluk analizi, veriler frekanslar ya da sayım verilerinin oluşturduğu tablolar biçiminde ifade edildiğinde hem sayısal hem de çoğunlukla grafiksel bir yöntem olarak kullanılır. Bu analiz, araştırmacıların çapraz tablodaki değişkenler arasındaki ilişkiyi anlama ve yorumlamasına olanak sağlar. Basit uygunluk analizi, çapraz tabloda yer alan iki değişken arasındaki ilişkiyi araştırır. Ancak uygulamalar, çoğunlukla ikiden fazla sayıda değişken içerir ve bu değişkenler arasındaki ilişkinin araştırılmasında çoklu uygunluk analizi kullanılır. Bileşik uygunluk analizi, çoklu uygunluk analizinin geliştirilmiş bir biçimi olarak ileri sürülmüştür. Bu çalışmanın temel amacı, Türkiye'deki yaşam memnuniyetini bileşik uygunluk analizi ve çeşitli çoklu uygunluk analizi yöntemlerini kullanarak incelemek ve yorumlamaktır.

Anahtar Kelimeler: Bileşik uygunluk analizi, Çoklu uygunluk analizi, Yaşam memnuniyeti araştırması

Çoker E. (2024). Joint Correspondence Analysis: A Case Study on Life Satisfaction in Türkiye. 19 Mayıs Sosyal Bilimler Dergisi, 5(3), 206-214. (Araştırma Makalesi)

Introduction

The aim of statistical research is often to try understand the relationships between the variables. In research including categorical data, the relationships among these variables are typically assessed using chi-square analysis. However, if the number of categories for row and column variables is large and their frequencies are insufficient, correspondence analysis is used to investigate these associations rather than chi-square analysis.

Correspondence analysis is a descriptive statistical technique that employs a geometric approach to represent a complicated categorical dataset in a comprehensible, low-dimensional (often two-dimensional) space. This method reveals the relationships between variables in the rows and columns of a very large-dimensional matrix while providing a visual representation of the raw data set.

Correspondence analysis resembles principal component analysis. The fundamental distinctions between both methodologies are that correspondence analysis is employed for categorical data, whereas principal component analysis is utilized for continuous data (Lebart et al., 1984). A further distinction between these methods is the approach by which the data matrix is partitioned. The principal component analysis decomposes the total variance, whereas the chi-square statistic is decomposed in correspondence analysis. In summary, correspondence analysis is based on mapping chi-square distances to Euclidean distances that are plotted in a graph as points. As the distance between these points decreases, the associated categories are interpreted as similar to each other. As the existence of a model is not necessary for the use of correspondence analysis, this method is often regarded as an exploratory data analysis technique (Greenacre, M. J., & Blasius, J., 2006).

Correspondence analysis is named in two different ways depending on the number of variables in the contingency table. The basic form of correspondence analysis, “Simple Correspondence Analysis”, is employed when dealing with two variables. "Multiple Correspondence Analysis" (MCA) is utilized when dealing with more than two variables.

Multiple Correspondence Analysis

MCA, a generalization of simple correspondence analysis to multivariate form, analyzes the relationship between categories of variables in the presence of more than two categorical variables. There are two different approaches for this analysis. In the first approach, all data are encoded as indicator variables, and the analysis is conducted using the indicator matrix. The second strategy involves evaluating all two-way cross-tabulations between variables via the Burt matrix.

The rows of the indicator matrix (Z) contain units, whereas columns contain categories of variables. The elements of this matrix consist of 1s and 0s, where a unit belongs to the relevant category of the variable; it is encoded as 1, otherwise as 0. Hence, the number of 1s in a row of the indicator matrix displays the number of variables. Consequently, the number of rows in the indicator matrix is the total number of units, and the number of columns is the total number of categories of all variables (Rencher, 2002).

The primary goal of MCA is to determine how much of the total variance (total inertia) is explained by the dimensions. The total variance of an indicator matrix depends only on the number of variables and the number of relevant categories in these variables, not on the original raw data. In MCA, separate indicator variables are created for each variable's categories. Assuming that Q is the total number of variables; J_q, number of categories in a variable (where q = 1, …, Q), so the total number of categories in all variables J can be obtained as follows:

$$J = \sum_{q=1}^{Q} J_q \tag{1}$$

For Q number of variables, the Z matrix is formed by placing the submatrices side by side:

$$Z = [Z_1, Z_2, \ldots, Z_q] \tag{2}$$

The total variance of the Z matrix is equal to the average of the variances of the submatrices and is calculated as follows:

$$\text{Variance(Z)} = \frac{1}{Q}\sum_{q=1}^{Q} inertia(Z_q) = \frac{1}{Q}\sum_{q=1}^{Q}(J_q - 1) = \frac{J-Q}{Q} \tag{3}$$

Since the dimension of the Z matrix is J-Q, the average variance for each dimension is 1/Q (Greeancre, 1992). The 1/Q value serves as a threshold to determine which dimensions require adequate interpretation in MCA, analogous to the eigenvalue threshold of "1" employed in principal component analysis.

The Burt matrix is another alternative data structure utilized in MCA, constructed from all pairwise relationships among the variables. The Burt matrix, denoted as the C matrix, is directly derived from the Z indicator matrix. This matrix consists of a diagonal containing Q variables crossed with themselves, along with Q(Q-1)/2 cross tables formed by pairs of variables, arranged symmetrically above and below the diagonal. This matrix is a symmetric square matrix (Greenacre, 2010).

The C matrix is obtained by multiplying the Z matrix from the left by its transpose:

$$C = Z^T Z \tag{4}$$

In the case of two variables, the C matrix is defined as follows:

$$Z^{'}Z = \begin{bmatrix} Z_1^{'}Z_1 & Z_1^{'}Z_2 \\ Z_2^{'}Z_1 & Z_2^{'}Z_2 \end{bmatrix} \tag{5}$$

This matrix is given below in case of Q number of variables:

$$C = \begin{bmatrix} Z_1^{'}Z_1 & Z_1^{'}Z_2 & \dots & Z_1^{'}Z_Q \\ Z_2^{'}Z_1 & Z_2^{'}Z_2 & \dots & Z_2^{'}Z_Q \\ \dots & \dots & \dots & \dots \\ Z_Q^{'}Z_1 & Z_Q^{'}Z_2 & \dots & Z_Q^{'}Z_Q \end{bmatrix} \tag{6}$$

Joint Correspondence Analysis

While MCA can be considered as a form of principal components analysis, joint correspondence analysis (JCA) can be considered as a factor analysis method applied to categorical variables (Vermunt & Andreson, 2005).

The Burt matrix version of MCA clearly demonstrates that the main problem in the analysis is to visualize the entire matrix. Hence, in this matrix, the main focus is on the off-diagonal elements. This method avoids diagonal elements and concentrates solely on off-diagonal elements, boosting overall variation while presenting more comprehensible visuals on maps (Greenacre, 2017).

JCA can be performed in a five-step method:

1) Greenacre's weighted least squares method is applied to the C matrix to obtain the C^*matrix.

$$C^* = \{c_{ij}^*\} = c_{ij} \tag{7}$$

2) MCA is performed using this C^*matrix.
3) The $\hat{C}$ matrix is formed by substituting the values estimated from the simple correspondence analysis with the original values in the diagonal. The following formula is used for this:

$$\hat{C} = \{\hat{n_{ij}}\} = nr_i r_j\left(1 + \sum_{s=1}^{S^*} \lambda_s a_{is} a_{js}\right) \tag{8}$$

In equation (8); a_{is}, is the i^{th} row (or column) standard coordinates for the s^{th} dimension; $u_{i1}$and u_{i2}, are the eigenvectors for the first two dimensions; r_i, is the row (or column) ratios; and $\lambda_s$is the basic variance.

4) Using equation (8), the main diagonal submatrices of C^*matrix are updated with the corresponding values in the $\hat{C}$ matrix.
5) Step 2 to step 4 is repeated until convergence is achieved.

Application: Türkiye Life Satisfaction Survey

This section of the study aims to apply JCA and various alternative MCA methods to data from the "Türkiye's Life Satisfaction Survey" performed by the Turkish Statistical Institute (TSI) in 2016. The purpose is to reveal, evaluate, and comprehend the overall profile of the data.

The "Life Satisfaction Survey" (LSS) was initially conducted in Türkiye in 2003 as part of the "Household Budget Survey" and then became an independent study in 2004. Provincial utilization of LSS commenced in 2013 (Beşel et al., 2015). TSI defines the objective of LSS as “to measure the individual's overall perception of happiness, social values, general satisfaction in fundamental living domains, and satisfaction with public services, as well as to track the fluctuations in these satisfaction levels over time.”

Since ancient times, people have addressed happiness and life satisfaction in various dimensions throughout the world and continue to do so today. The concept of happiness has been put forward in various aspects over the years, from the subjective appreciation of individuals with their material or spiritual values to having values that will gain the appreciation of society, to their combinations in various proportions, to the satisfaction with the way of life they lead (TÜİK, 2016). However, today's approach to happiness has evolved to focus on whether an individual feels happy or not, and this definition has gained significant importance.

The primary objective of the study is to assess the levels of satisfaction and happiness individuals experience in LSS. This study covers Turkish citizens and foreign nationals aged 18 and over living in households within the borders of the Republic of Türkiye. The survey excluded institutional populations, such as university dorms, nursing homes, jails, detention facilities, reformatories, orphanages, and military units or barracks.

The LSS, performed in 2016, involves 8,981 participants. Since there are a large number of variables in LSS, seven variables associated with happiness and satisfaction are included in the scope of this study, as detailed follows: The “gender” variable is encoded as female and male; the “age” variable is defined as four categories: 18-29, 30-49, 50-64, and 65 and above. There are five categories in the "happiness—How happy are you when you think about your life as a whole?” variable: very happy, happy, moderate, unhappy, very unhappy. These categories are grouped under three categories: “happy, moderate, and unhappy. For the “next year—How do you think your life will be in general next year?” variable, it is expressed under four categories: improve, same, worse, and no idea. Regarding the variable “future comparison—When you think about the next five years, how do you estimate your situation will be in general?”, there are four categories in total: improve, same, regress, and no idea. For the "hope—Are you hopeful about your own future?” variable, there are four categories: very hopeful, hopeful, not hopeful, not hopeful at all. These categories are grouped into two categories: hopeful and hopeless. There are six categories for the variable “what happy—What makes you happiest in life?" encompasses six categories: power, work, love, success, health, and money. This variable is combined into four categories: career, health, love, and money.

The frequency table for all variables used in the analyses conducted in the study is given in Table 1:

Table 1. Frequency table of variables used in the analysis

Variables	Categories	Frequencies (%)	Variables	Categories	Frequencies (%)
Gender	Male	4035 (% 45)	Future Comparison	Improve	3640 (41%)
	Woman	4946 (% 55)		Same	2865 (32%)
Age	18-29	1879 (% 21)		Regress	1467 (16%)
	30-49	3731 (% 41)		No idea	1009 (11%)
	50-64	2146 (% 24)	Hope	Hopeful	6831 (76%)
	65+	1225 (14%)		Hopeless	2150 (24%)
Happiness	Happy	5491 (61%)	What Makes You Happy	Career	766 (8%)
	Moderate	2592 (29%)		Health	6549 (73%)
	Unhappy	898 (10%)		Love	1321 (15%)
Next Year	Improve	3669 (41%)		Money	345 (%4)
	Same	3552 (39%)			
	Worse	875 (10%)			
	No idea	885 (10%)			

Application of Multiple Correspondence Analysis

All applications carried out within the study are carried out using the ca (Nenadic & Greenacre, 2007) and FactoMineR (Lê, Josse & Husson, 2008) packages in the R program (R Core Team, 2021). Three variations of MCA applications of LSS are carried out with the indicator matrix, Burt matrix and adjusted Burt matrix.

Multiple Correspondence Analysis with Indicator Matrix

The map obtained as a result of the MCA applied using the indicator matrix are given below:

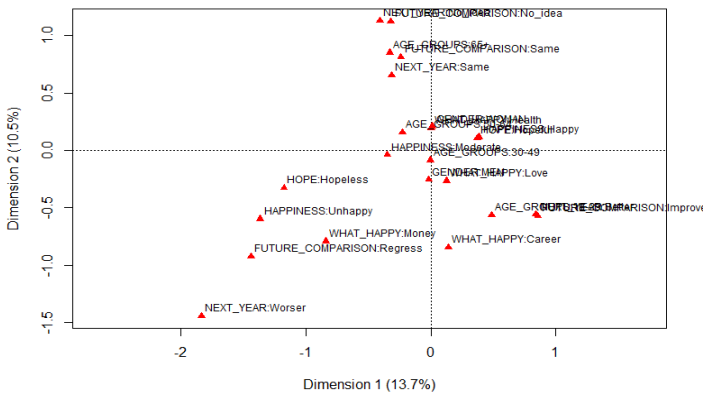

Figure 1. MCA map (indicator matrix based)

When examining Figure 1, it is observed that the first two dimensions in the MCA, derived from the indicator matrix, account for 24% of the variation. The inclusion of categories with variable names in Figure 1 leads to the map more complex. Consequently, the map, which excludes variable names and incorporates just the relevant categories of the variables, is organized and shown below:

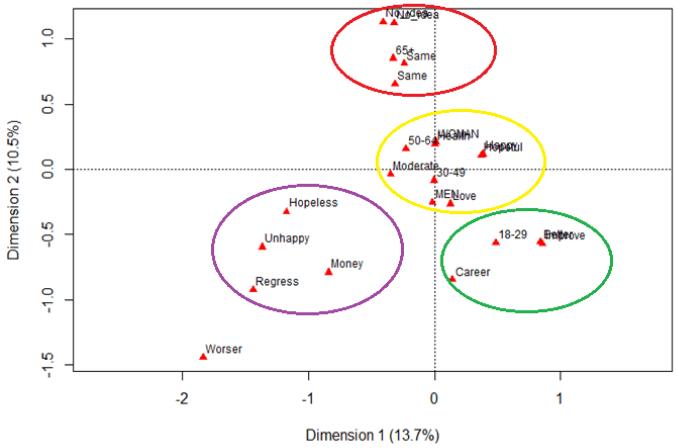

Figure 2. Modified MCA map (indicator matrix based)

A closer look at the modified map in Figure 2 reveals that variable categories that are near to one another are circled in various colors to form a total of four groupings. Individuals who identify themselves as unhappy have a sense of hopelessness, believing that their lives will get worse over the next five years, with the primary source of their happiness being financial wealth. Individuals 65 years of age and older make up a different group. Individuals in this group have expressed uncertainty regarding their life in the next year and five years, or that they would remain unchanged. The third group comprises those aged 18-29 who anticipate an improvement in their lives over the next year and in five years, attributing their greatest happiness to their careers. The fourth and last group comprises those aged 50-64, who report that love is the primary source of their happiness and express a general sense of hopefulness. It is evident from examining the map that gender has no discernible impact because it is nearly on the axes.

Multiple Correspondence Analysis with Burt Matrix

The map obtained as a result of the MCA using the Burt matrix are as follows:

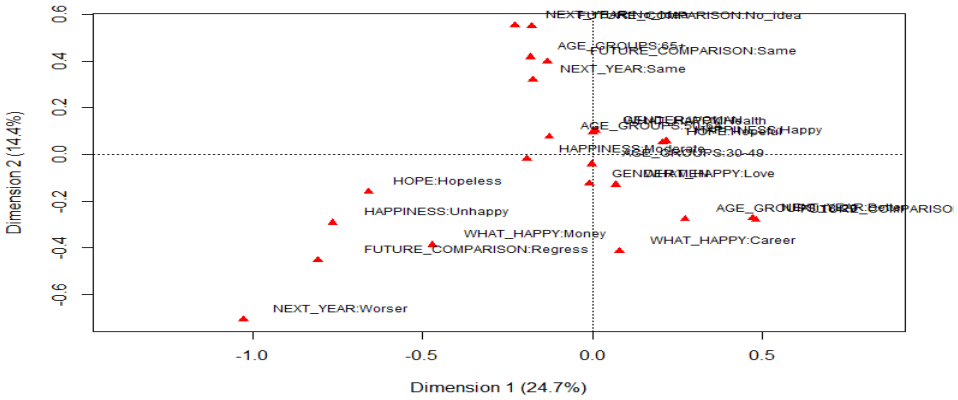

Figure 3. MCA map (Burt matrix based)

By examining Figure 3, it is observed that the first two dimensions in the MCA derived from the Burt matrix account for 39% of the variation. Similar to Figure 1, the map in Figure 3 appears to be complicated because the categories are displayed with the variable names. Consequently, the map, from which variable names have been removed and only the relevant categories of the variables are presented, is organized and given below: Similar to the indicator matrix approach, gender exhibits no influence, since it is almost aligned with the axes.

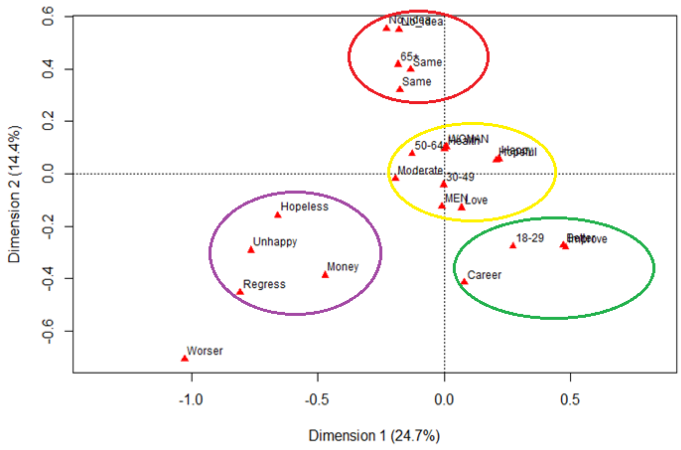

Figure 4. Modified MCA map (Burt matrix based)

The profiles depicted in Figure 4 are identical with Figure 2. The only distinction between the indicator matrix and the Burt matrix approaches is the explanation rate. The explanation rate increases from 24% to 39% in the Burt matrix results.

Multiple Correspondence Analysis with Adjusted Burt Matrix

The adjusted Burt matrix explains the Burt matrix together with Greenacre adjusted total variance values. Therefore, a higher explanation rate is expected when MCA is performed with an adjusted Burt matrix. The maps generated by the application of MCA using the updated Burt matrix for the LSS data are presented below:

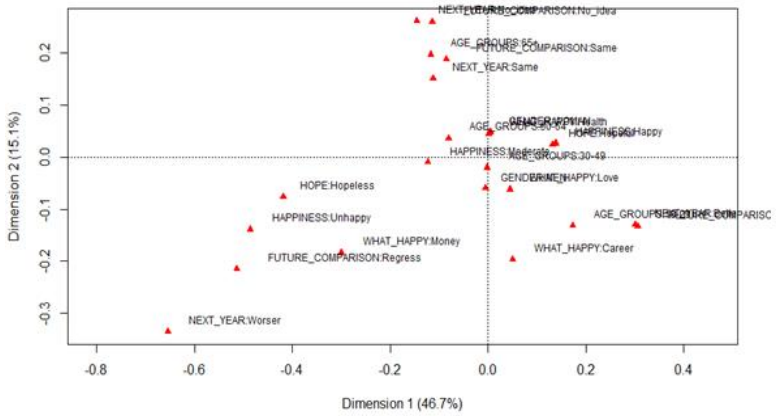

Figure 5. MCA map (Adjusted Burt matrix based)

Figure 5 reveals that 62% of the variation can be explained by the first two dimensions in the MCA implemented using an adjusted Burt matrix. Figure 6 below presents the map with the variable names excluded, retaining just the relevant categories of the variables from Figure 5.

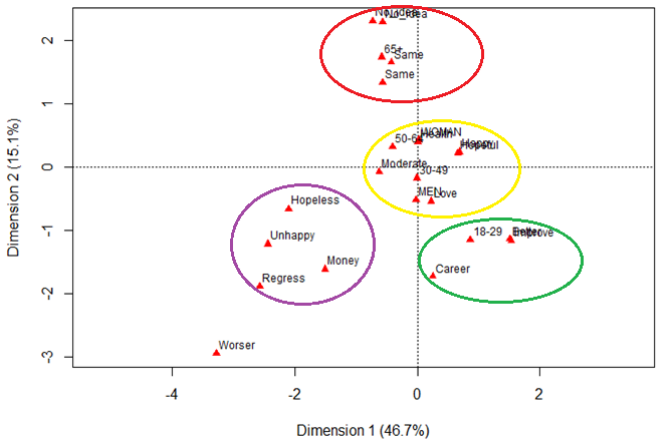

Figure 6. Modified MCA map (Adjusted Burt matrix based)

Examining Figure 6 reveals that the remarks expressed in Figures 2 and 4 are also applicable here.

Application of Joint Correspondence Analysis

The map of the JCA is obtained as follows:

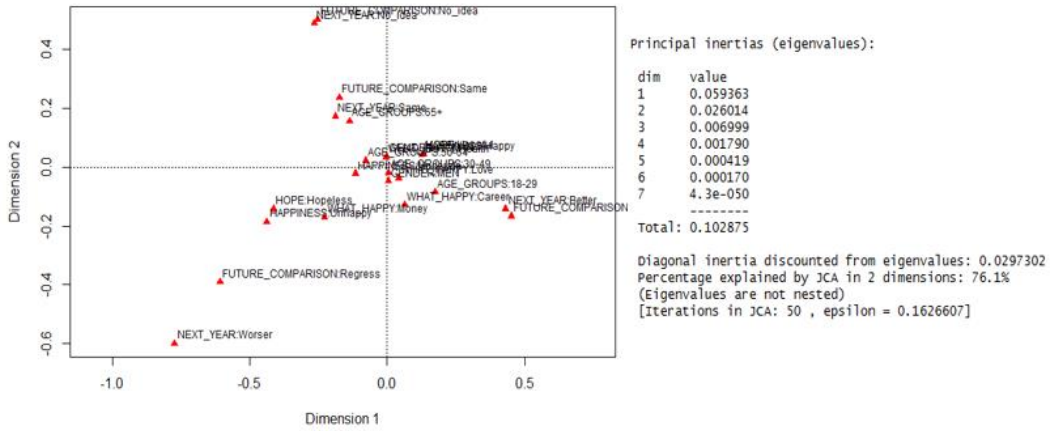

Figure 7. JCA map

Analysis of Figure 7 indicates that the first two dimensions account for 76% of the variation when JCA is utilized. The revised map, Figure 8, which presents the map with the variable names removed and only the relevant categories included, is shown below:

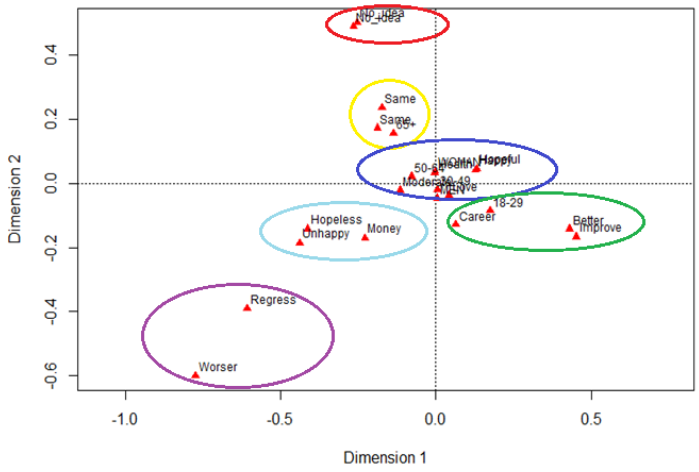

Figure 8. The final JCA map

If Figure 8 is evaluated, it is understood that the groupings obtained in the MCA are generally the same here, but only two points stand out. According to the map in Figure 8, the individuals categorized in a single group by all variations of MCA were divided into two distinct groups: those who expressed uncertainty regarding their life circumstances in the upcoming year and in five years, and those aged over 65 who anticipated that their life situation would remain unchanged both in the next and five-year time line. A remarkable observation is that those who are hopeless and claim that their primary source of happiness is wealth create a distinct group, while those who anticipate a decline in their quality of life over the next year and in five years represent another unique group.

The table below presents the explanation rates obtained from the comparison of the four distinct methodologies.

Table 2. Explanation Rates Obtained from the Methods Used in the Analysis

Method	**Explanation Ratio**
MCA with Indicator Matrix	24%
MCA with Burt Matrix	39%
MCA with Adjusted Burt Matrix	62%
JCA	76%

Table 2 indicates that the JCA possesses the maximum explanatory power, as theoretically anticipated. This scenario reaffirms the primary rationale for employing JCA via the LSS data.

Discussion and Conclusion

This study applied various correspondence analysis techniques, including MCA with indicator, the Burt and adjusted Burt matrices, as well as JCA, to Türkiye's 2016 Life Satisfaction Survey (LSS). These methods aimed to explore the relationships among categorical variables related to life satisfaction, happiness, and demographic characteristics. The findings provide valuable insights into both the data and the efficacy of the analytical methodologies.

The research illustrates that MCA and JCA proficiently capture and represent relationships among categorical variables, making them especially appropriate for survey data such as the LSS. Three principal observations arose from the analyses. The initial section relates to demographic and psychological patterns. Younger individuals (18-29 years old) typically expressed contentment and hopefulness towards career fulfillment and maintained a positive outlook on their future. Individuals aged 65 and older were more inclined to express stability or uncertainty about their future, constituting a separate cohort. The prioritization of material wealth as a source of happiness significantly correlated with hopelessness.

The next section concerns the efficacy of the analytical techniques. The MCA utilizing the adjusted Burt matrix accounted for 62% of the variation, whereas the JCA attained the best explanatory power at 76%. JCA provided a more sophisticated categorization of persons than MCA. For instance, it categorized older persons into those expecting stable conditions and those indicating doubt regarding their future. This nuanced difference illustrates JCA's capability to handle complex interactions among variables effectively.

The final section addresses the small influence of gender. Across all methodologies, gender exhibited minimal impact, closely correlating with the axes, indicating it had a limited function in influencing respondents' judgments of happiness and contentment.

This study highlights the effectiveness of JCA as a powerful exploratory instrument for examining categorical survey data. Its enhanced explanatory capacity relative to MCA makes it especially beneficial for determining complex relationships and classifications. Within the framework of Türkiye's LSS, JCA indicated complex insights into life satisfaction patterns, including the unique profiles of individuals categorized by age, happiness, and sources of hope or dissatisfaction.

Incorporating longitudinal data or broadening the variables could enhance the analysis in future studies and offer a more dynamic picture of patterns in life satisfaction across time.

This study illustrates the methodological improvements of JCA offers a significant interpretation of life satisfaction in Türkiye, therefore contributing to academic literature and informing policymaking aimed at enhancing social well-being.

References

Beşel, F., Yardımcıoğlu, F., Gürdal, T. (2015). Yaşam Memnuniyeti Araştırması Kapsamında

Çalışma Hayatının Analizi: 2003-2014 Dönemi, Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi, 16. Çalışma Ekonomisi ve Endüstri İlişkileri Kongresi Özel Sayısı.

Clausen, S –E. (1998). Applied Correspondence Analysis-An Introduction, Sage Publication,London.

Fisher, R. A. (1940). The precision of discriminant functions. Annals of Eugenics, 10,422-429.

Greenacre, M. (2017). Correspondence Analysis in Practice, Third Edition, CRC Press.

Greenacre M. (2010). Biplots in Practice, Fundacion BBVA, Spain.

Greenacre, M. J., & Blasius, J. (2006). Multiple correspondence analysis and related methods. Chapman & Hall/CRC.

Greenacre, M. (1992). Correspondence Analysis in Practice, Chapman & Hall/CRC Press,New York.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (with collaboration of P. Wallin & L. Guttman) (Ed.), The prediction of personal adjustment, 319-348, New York: Social Science Research Council.

Hirschfield H.O. (1935). A connection between correlation and contingency. Proc Camb Phil Soc. 1935; 31:520–4.

Kaptan, Y. (2010). Sağlık sektöründe kalite iyileştirmesi ve istatistiksel yöntemlerle incelenmesi, Yüksek lisans tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Lê S., Josse, J. & Husson, F. (2008). "FactoMineR: A Package for Multivariate Analysis." Journal of Statistical Software, **25**(1), 1–18.

Lebart, L., Morineau, A. & Warwick, K. (1984). Multivariate Descriptive Statistical Analysis.Chichester, UK: Wiley.

Nenadic, O. & Greenacre, M. (2007). "Correspondence Analysis in R, with two- and three-dimensional graphics: The ca package." Journal of Statistical Software, **20**(3), 1-13.

R Core Team, (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rencher A.C. (2002). Methods of Multivariate Analysis, A John Wiley & Sons, Inc., Canada.

Suner, A., Çelikoğlu, C.C. (2008). Uygunluk Analizinin Benzer Çok Değişkenli Analiz Yöntemleri İle Karşılaştırılması, İstatistikçiler Dergisi: İstatistik ve Aktüerya, 1: 9-15.

Suner, A. (2007). Application of a population based study of correspondence analysis in choosing a health institution, PhD Thesis, Dokuz Eylül Üniversitesi Graduate School of Natural and Applied Sciences, İzmir.

Türkiye İstatistik Kurumu, (2016). Yaşam Memnuniyeti Araştırması, Türkiye İstatistik Kurumu, Ankara.

Vermunt J.K. and Anderson C.J. (2005). Joint Correspondence Analysis (JCA) by Maximum Likelihood, Department of Methodology and Statistics, Tilburg University, The Netherlands.